DOCUMENT RESUME

ED 286 883 TM 870 135

AUTHOR Avant, Anna H.; O'Neal, Marcia R.
TITLE Investigation of the Otis-Lennon School Ability Test to Predict WISC-R Full Scale IQ for Referred Children.
PUB DATE Nov 86
NOTE 15p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (15th, Memphis, TN, November 19-21, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Correlation; Diagnostic Tests; Elementary Secondary Education; Gifted; *Intelligence Tests; Learning Disabilities; Mild Mental Retardation; Predictive Measurement; *Racial Differences; Sex Differences; *Student Placement; *Test Bias
IDENTIFIERS *Otis Lennon School Ability Test; Potthoff Technique; *Wechsler Intelligence Scale for Children (Revised)

ABSTRACT

A public school district in Alabama uses the Otis Lennon School Ability Test (OLSAT) in place of the Wechsler Intelligence Scale for Children (Revised) (WISC-R) as a screening device for students who are referred for possible special education placement. To determine the relationship between the two tests, Pearson product moment correlation coefficients were computed for all referrals and for the following subgroups: (1) sex; (2) race; and (3) subsequent placement category (gifted, learning disabled, EMR, and no recommended placement). Also, the Potthoff technique was used to detect bias for sex, race, or the total group. WISC-R Full Scale IQ scores, OLSAT scores, and OLSAT School Ability Index were available for 431 of the students referred for placement in 1985-86. Subjects were ages 6-16; 283 were male, 148 female; 273 were white, 158 black. Correlations between the two tests were statistically significant for all but the gifted group. Correlations were slightly higher for females and whites; the EMR subgroup showed the lowest correlation. The Potthoff technique revealed a bias in prediction for race, but not for sex, for the group as a whole. Differences were also revealed within the learning disabled group and those not recommended for placement. (MGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED286883

Investigation of the Otis-Lennon School Ability Test to Predict WISC-R Full Scale IQ for Referred Children

Anna H. Avant
Marcia R. O'Neal

The University of Alabama

Paper Presented at the Fifteenth Annual Meeting of the
Mid-South Educational Research Association
Memphis, Tennessee
November 19-21, 1986

BEST COPY AVAILABLE

TM 870 135

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. H. Avant

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

INVESTIGATION OF THE OTIS-LENNON SCHOOL ABILITY TEST TO PREDICT WISC-R FULL SCALE IQ FOR REFERRED CHILDREN

Anna H. Avant, Marcia R. O'Neal, University of Alabama

The Full Scale IQ (FSIQ) measured by the Wechsler Intelligence Scale for Children-Revised (WISC-R) is widely used as one element of information contributing to placement decisions. Because the WISC-R requires individual administration by a highly trained evaluator, concern for efficient use of the evaluator's and the student's time emphasizes the need for a screening measure which will minimize unnecessary testing while identifying those students who are candidates for further evaluation and possible placement.

An instrument currently in use in one southern Alabama school system is the Otis-Lennon School Ability Test (OLSAT). Within this system, the OLSAT is routinely administered in conjunction with the Stanford Achievement Test (SAT) and yields a School Ability Index (SAI) with a mean of 100 and standard deviation of 16. The most recent version of the OLSAT, which is a revision of the Otis-Lennon Mental Ability Test (OLMAT), was standardized concurrently with the SAT and was published in 1982. Correlations of the OLSAT with the OLMAT range from .76 to .94. Internal consistency reliability ranges from .92 to .96, and alternate form reliabilities are reportedly between .82 and .92. Test-retest reliabilities vary between .84 and .92. The use of the OLSAT as a referral screening device represents the potential for a substantial savings of financial and personnel resources as well as improved service to students.

The purpose of this study was to provide empirical evidence on which to judge the OLSAT as a screening device by demonstrating its relationship to the WISC-R FSIQ for students referred for evaluation. Pearson product moment correlation coefficients were computed for all referrals as well as for subgroups by sex, race, and ultimate placement category (gifted, LD, EMR, and no recommended placement). The Potthoff technique was also employed to detect the presence of sex or race bias for the total group and for each placement category. This technique yields an omnibus F ratio which allows for the simultaneous testing of the constancy of both the regression coefficients and intercept values across the groups being compared. Detection of bias would be the basis for the development of separate prediction equations for the groups.

The school system in the study consists of both urban and rural areas and includes inner city schools. A total of 66,000 students are served in the system's 91 schools. The county served, extending 1,248 square miles, has a population of 364,379. The per capita income of the county is $17,011. Students referred during the 1985-86 school year ranged in age from 6 to 16 years and included 283 males and 148 females; 273 were white and 158 were black. Of the total, 23 were placed in gifted programs, 158 in programs for learning disabilities, and 38 EMR placement. The remaining 212 were not placed. The SAI scores ranged from 50 to 143 with a mean of 90.02 and standard deviation of 19.80. WISC-R FSIQs ranged from 45 to 141 with a mean of 91.40 and standard deviation of 17.02. The SAI-FSIQ correlation for the total group was .82, with correlations for the subgroups ranging from -.11 to .87. The lowest correlations (from -.11 to .37) were for the gifted placements while the highest correlations (.73 to .86) were for the nonplaced category.

Results of the Potthoff technique revealed bias in prediction based on race for the total group and for the LD and nonplaced categories. Where bias was demonstrated, prediction equations were calculated for the separate groups. Implications of results for the school system were discussed.

Investigation of the Otis-Lennon School Ability Test to Predict WISC-R Full Scale IQ for Referred Children

Anna H. Avant and Marcia R. O'Neal
The University of Alabama

Introduction and Background

In the 1980s there has been increased interest in identifying children eligible for special education programs. Due to the passage of P.L. 94-142 and heightened public awareness of such programs, there has been an upsurge in the number of referrals for psychological evaluations to determine eligibility for special education placement. Because of this trend, it has become increasingly important to develop accurate screening measures which will effectively limit the number of students who will be evaluated. This is particularly true in light of the current climate of limited funding.

At the present time, the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) is the most prevalently used instrument for determining the eligibility of students for programs for the mentally retarded, learning disabled, and gifted. The Otis-Lennon School Ability Test (OLSAT), a group test of school learning ability, is one of the most frequently used group intelligence tests in the U.S. It is often used to screen children being considered for placement in special education classes.

Because it is very probable that children referred for such classes on the basis of their OLSAT scores will be further evaluated with the WISC-R, it is imperative that the two

instruments be compared. Validity studies have indicated that WISC-R Full Scale IQs of 88 average students ranging in age from 9 to 15 years correlated .85 with the OLSAT. In a smaller study, the WISC-R Full Scale IQ correlated .60 to .75 with Form R of the OLSAT across several levels (Dyer, 1985). Dyer indicated that the results of these studies were of limited value because of the small sample sizes and the selectiveness of the samples. The purpose of the present study was to provide additional empirical evidence on which to judge the OLSAT as a screening device by demonstrating its relationship to the WISC-R Full Scale IQ for students referred for evaluation.

Method

Subjects and Procedure

The subjects who comprised the sample were children referred for evaluation during the 1985-86 school year in a public school district in southern Alabama. The county served by this school system extends 1,248 square miles. Its population is 364,379, per capita income is $17,011, and there are 91 schools in the system attended by 68,000 students.

The records of those students evaluated during the current school year were searched. The WISC-R Full Scale IQ of each student whose records indicated prior OLSAT scores was recorded, as was each student's OLSAT School Ability Index (SAI). All students whose records included both tests were selected for inclusion in the study. Certified school psychologists and psychometrists with extensive experience in administering and scoring individual psychoeducational tests to subjects of various ages and cultures administered all WISC-R evaluations.

Subjects included in the study ranged in age from 6 to 16 years. The sample included 283 males and 148 females. Of the total of 431 students, 273 students were white and 158 were black. Ultimate placement categories included 23 recommended for placement in the gifted program, 158 for placement in a program for the learning disabled, and 38 for EMR placement. Two hundred twelve received no recommendation for placement.

Materials

The Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) is an individually administered intelligence test appropriate for children and youth whose ages range from 6 to 16 years. It is comprised of two separate scales. The Verbal Scale consists of the Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span (optional) subtests. The Performance Scale is composed of the Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Mazes (optional) subtests. Verbal, Performance, and Full Scale IQs are computed, each having a mean of 100 and a standard deviation of 15. This test is administered in approximately 50 to 75 minutes.

Form S of the Otis-Lennon School Ability Test (OLSAT) is a revision of the Otis-Lennon Mental Ability Test (OLMAT). Published in 1982, it was standardized concurrently with the Stanford Achievement Tests on a representative sample of over 130,000 students (Dyer, 1985). As Oakland (1985) indicated, "The theoretical basis and purpose of the OLSAT are similar to those of the earlier tests", incorporating "Spearman's theory (1927) as modified by Vernon (1960) [which] conceptualizes intelligence as two general factors, a verbal-educational factor and a

practical-mechanical factor" (p. 1111). Dyer stated that "the test was designed to assess only the verbal-educational domain" and that "Guilford's Structure of Intellect view was used as a taxonomy for selecting items which would be balanced for the intellectual operations of cognition, convergent thinking, and evaluation" (p. 1107). He further noted that the OLSAT "gives a single score and retains an omnibus format with a mixture of verbal, numerical, and figural items in spiral order" (p. 1107).

The single score derived from the OLSAT has a mean of 100 and standard deviation of 16, making it easily related to IQ measures. It is, however, reported as a School Ability Index (SAI) in order to emphasize abilities learned by experience and schooling rather than those innate abilities and to reduce the risk of misinterpretation (Dyer, 1985).

Ahmann (1985) indicated that the OLSAT correlated well with its predecessor, the OLMAT, suggesting the similarity between the two instruments. Coefficients ranged from a low of .76 for first graders to a high of .94 for students in the 11th grade. Dyer (1985) discussed similar coefficients found between the two forms (Form S and Form R) of the OLSAT. Correlations of .82 at the Primary I level to .92 at the Intermediate level were found.

Internal consistency reliability coefficients for Form S using the Kuder-Richardson Formula 20 ranged from .92 to .96, according to Ahmann (1985). Oakland (1985), in his review of the test, indicated coefficients for age which ranged from .90 to .95 and coefficients for grade which ranged from .88 to .94. These values, he contended, provide ample evidence for the OLSAT's internal consistency and homogeneity.

Evidence for the stability of the OLSAT was found, according to Oakland (1985), in correlations from students at five grade levels who were retested after 6 months. These correlations ranged from .84 to .92 for Form R. As Dyer (1985) stated, however, test-retest reliabilities are not available in the manual for Form S.

Validity evidence for the OLSAT is less extensive. Dyer's (1985) review indicated correlations of the WISC-R Full Scale IQ with Form R ranging from .60 to .75 across several levels, and in a larger (N=88) sample of average students, correlations ranged from .85 to .88. Correlation of Form R with the Stanford Binet Intelligence Scale for a sample of 184 above average students was .58. Oakland (1985) indicated that the correlation of the OLSAT with the WISC-R Full Scale IQ for grades 3 through 9 was .85. Despite limited validity information, Dyer (1985) was willing to conclude the following:

> The O-LSAT Forms R and S show psychometric sophistication and rigor in standardization consistent with the past edition of the O-LMAT. Evidence is positive for reliabilities of the test based on internal consistency of items, stability of retesting, and equivalence of the two forms. (p. 1110)

Analysis of Data

Pearson product moment correlation coefficients were computed for all referrals, as well as for subgroups by sex, race, and subsequent placement category (gifted, LD, EMR, and no recommended placement). The Potthoff technique was also employed to detect the presence of sex or race bias for the total group and for each placement category. This technique is a statistical method for examining external criteria which determine whether the regression

equation relating any two variables is constant across groups. This procedure allows the simultaneous testing of the equivalence of regression coefficients and intercepts across a given number of independent groups with a single overall F ratio. A significant omnibus F allows for the separate testing of slopes and intercepts (Reynolds, 1982). Detection of bias would be the basis for the development of separate prediction equations for the various groups in order to achieve fairness in prediction.

Results and Discussion

SAI scores ranged from 50 to 143, while WISC-R Full Scale IQs ranged from 45 to 141 for the students included in the study. Means, standard deviations, and SAI - WISC-R correlations for the total sample and for subgroups by sex, race, and placement group are shown in Table 1.

Table 1

Means, Standard Deviations, and Correlations for WISC-R Full Scale IQs and SAI Scores for Race, Sex, and Placement Subgroups

		SAI		WISC-R		
Subgroup	N	$\bar{X}$	SD	$\bar{X}$	SD	r
Total	431	90.02	19.80	91.40	17.02	.82**
Males	283	88.96	18.34	90.98	16.09	.79**
Females	148	92.03	22.25	92.19	18.69	.87**
Whites	273	95.01	18.79	96.66	16.18	.83**
Blacks	158	81.40	18.54	82.30	14.42	.74**
Gifted	23	126.52	8.10	127.43	6.33	.10
LD	158	88.66	13.76	92.92	10.08	.57**
EMR	38	64.82	9.53	66.18	6.58	.42**
No Rcmded Plcmt	212	91.58	19.28	90.87	15.64	.81**

** $p < .01$

All correlations, with the exception of the gifted placement group, were statistically significant ($p < .01$). Correlations were slightly higher for females than for males, and were higher for whites than for blacks. Within the placement subgroups, the lowest significant correlation was for EMR students ($r=.42$). For this group, less than 20% of the variance was shared ($r2=.18$).

Separate correlations by sex and by race within each placement category revealed consistently lower correlations for blacks and for males. Within the gifted category, the SAI - WISC-R correlation for black students was -.11, and for males it was .08. The highest correlations were for white students ($r=.84$) and females ($r=.86$) who were not placed. These and remaining subgroup correlations are shown in Table 2.

Table 2

SAI - WISC-R Correlations Within Each Placement Category

Group	N	r
Gifted		
Males	14	.08
Females	9	.37
Whites	20	.20
Blacks	3	-.11
LD		
Males	125	.55**
Females	33	.63**
Whites	110	.57**
Blacks	48	.40**
EMR		
Males	22	.37*
Females	16	.44*
Whites	10	.69*
Blacks	28	.35*
No Rcmded Plcmt		
Males	122	.76**
Females	90	.86**
Whites	133	.84**
Blacks	79	.73**

* $p < .05$ ** $p < .01$

It should be noted that several of the correlations are based on extremely small numbers; of particular note are those in the gifted and EMR categories, all of which are less than 30.

Results of the application of the Potthoff technique for detecting bias in prediction for sex and for race are reported in Table 3.

Table 3

Results of Potthoff Technique for Detecting Prediction Bias for Race and Sex

Group	Omnibus F	df	Slope F	df	Intercept F	df
Total						
Sex	.82	2,427	.69	1,427	.95	1,428
Race	18.31**	2,427	7.47**	1,427	28.71**	1,428
Gifted						
Sex	1.57	2,19	.09	1,19	3.21	1,20
Race	.40	2,19	.21	1,19	.61	1,20
LD						
Sex	.16	2,154	.18	1,154	.13	1,155
Race	11.14**	2,154	6.22*	1,154	15.53**	1,155
EMR						
Sex	.58	2,34	.22	1,34	.95	1,35
Race	.51	2,34	.44	1,34	.59	1,35
No Rcmded Plcmt						
Sex	1.53	2,208	2.81	1,208	.25	1,209
Race	12.42**	2,208	8.63**	1,208	15.63**	1,209

* $p < .05$ ** $p < .01$

For the students in the study, no prediction bias existed based on sex, but bias in prediction was present based on race. Both the slope and the intercept for the prediction equation differed for black and white students for the group as a whole and within the LD group and those for which no placement was recommended.

In practical terms, a statistically significant difference between intercepts alone would mean that the use of a common regression equation would underpredict scores for students in the group with a higher mean criterion score and would overpredict scores for students in the group with lower mean criterion score, a bias which would remain constant across score values. A statistically significant difference between slopes alone would also result in underprediction for the higher group, but the bias in prediction would vary depending on the distance of the score from the mean. A more complex situation exists if both slopes and intercepts are significantly different, as is the present case. The direction of bias may reverse and underprediction may sometimes result for students in the group with the lower mean criterion score, depending on the location of the predictor variable score (Reynolds, 1982). Thus, when prediction bias is detected, separate prediction equations should be developed for the groups.

The slopes and intercepts were computed for black and white students separately for groups in which prediction bias was indicated. Separate prediction equations were then developed. These values are given in Table 4.

Table 4

Slopes, Y-Intercepts, and Prediction Equations for Black and White Students

Group	Slope	Y-Intercept	Prediction Equation
Total			
Whites	.96	2.13	y=2.13+.96x
Blacks	.95	3.25	y=3.25+.95x
LD			
Whites	.74	20.56	y=20.56+.74x
Blacks	.82	11.68	y=11.68+.82x
No Rcmded Plcmt			
Whites	1.80	.49	y=.49+1.80x
Blacks	1.06	-3.69	y=-3.69+1.06x

Conclusion

While statistically significant correlations were found between the SAI and the WISC-R Full Scale IQ for this group of 431 referred students, it is clear that the relationships were not of equivalent strength for all groups. Further, the results of application of the Potthoff technique revealed that the SAI did not predict similarly for black and white students. School systems using the SAI as a screening device to predict WISC-R Full Scale IQs should be aware of this differential prediction for the two groups. The use of separate prediction equations for black and white students would, according to the present results, yield more accurate predictions of WISC-R Full Scale IQs. Of course, it should be pointed out that the data here represent the simple case in which only one variable contributes to the prediction of the WISC-R Full Scale IQ.

Ahmann, J. S. (1985). Otis-Lennon School Ability Test. Measurement and Evaluation in Counseling and Development, 17, 226-229.

Dyer, C. O. (1985). Review of Otis-Lennon School Ability Test. In J. V. Mitchell (Ed.), The Ninth Mental Measurements Yearbook (pp. 1107-1111). Lincoln, NE: Buros Institute of Mental Measurements.

Oakland, T. (1985). Review of Otis-Lennon School Ability Test. In J. V. Mitchell (Ed.), The Ninth Mental Measurements Yearbook (pp. 1111-1112). Lincoln, NE: Buros Institute of Mental Measurements. Buros 1111-1112.

Reynolds, C. R. (1982). Methods for detecting construct and predictive bias. In R. A. Berk (Ed.) Handbook of methods for decting test bias. Baltimore, MD: University Press.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children-Revised (WISC-R). New York: The Psychological Corporation.